Theor Appl Genet (2010) 121:845–856
DOI 10.1007/s00122-010-1354-1

ORIGINAL PAPER

Molecular characterization and comparative transcriptional analysis of LMW-m-type genes from wheat (*Triticum aestivum* L.) and *Aegilops* species

X. H. Li · K. Wang · S. L. Wang · L. Y. Gao · X. X. Xie · S. L. K. Hsam · F. J. Zeller · Y. M. Yan

Received: 9 February 2010 / Accepted: 1 May 2010 / Published online: 19 May 2010

Abstract Twelve new LMW-GS genes were characterized from bread wheat (*Triticum aestivum* L.) cultivar Zhongyou 9507 and five *Aegilops* species by AS-PCR. These genes belong to the LMW-m type and can be classified into two subclasses designated as 1 and 2, with the latter predominant in both wheat and related wild species. Genes in the two subclasses were significantly different from each other in SNPs and InDels variations. In comparison to subclass 1, the structural features of subclass 2 differs in possessing 21 amino acid residue substitutions, two fragment deletions (each with 7 amino acid residues), and a double-residue deletion and two fragment insertions (12 and 2–5 residues). Phylogenetic analysis revealed that the two subclasses were divergent at about 6.8 MYA, earlier than the divergence of C, M, N, S^s and U genomes. The S^s and B genomes displayed a very close relationship, whereas the C, M, N and U genomes appeared to be related to the D genome of bread wheat. The presently characterized genes *ZyLMW-m1* and *ZyLMW-m2* from Zhongyou 9507 were assigned to the D genome. Moreover, these genes were expressed successfully in *Escherichia coli*. Their transcriptional levels during grain developmental stages detected by quantitative real-time PCR (qRT-PCR) showed that both genes started to express at 5 days post-anthesis (DPA), reaching the maximum at 14 DPA after which their expressions decreased. Furthermore, the expression level of *ZyLMW-m2* genes was much higher than that of *ZyLMW-m1* during all grain developmental stages, suggesting that the expression efficiency of LMW-GS genes between the two subclasses was highly discrepant.

Communicated by X. Xia.

X. H. Li, K. Wang, and S. L. Wang contributed equally to this work.

X. H. Li · K. Wang · S. L. Wang · L. Y. Gao · X. X. Xie · Y. M. Yan (✉)
Key Laboratory of Genetics and Biotechnology, College of Life Science, Capital Normal University, Beijing 100048, China
e-mail: yanym@hotmail.com

S. L. K. Hsam (✉) · F. J. Zeller
Lehrstuhl für Pflanzenzüchtung, Technische Universität München, 85350 Freising-Weihenstephan, Germany
e-mail: hsam@wzw.tum.de

Introduction

Glutenins, the major determinant of bread-making quality in wheat endosperm, consist of high molecular weight glutenin subunits (HMW-GS) and low molecular weight glutenin subunits (LMW-GS), with profound effects on dough elasticity and viscosity, respectively. In the past decades, both subunits have been widely investigated for their genetic and biochemical characteristics (Gianibelli et al. 2001; Shewry and Halford 2002). Data on a considerable number of HMW-GS genes and more than 200 genomic DNA and cDNA clones of LMW-GS genes from hexaploid wheat and related species have been published (Yan et al. 2004; Huang and Cloutier 2008).

LMW-GS accounted for about 70% of the total amount of glutenins in endosperm and are encoded by *Glu-A3*, *Glu-B3* and *Glu-D3* loci on the short arms of chromosome 1A, 1B and 1D, respectively. In comparison to HMW-GS, it was difficult to separate LMW-GS and to identify their relationships with dough quality due to the complexity of this multigene family. Based on the first N-terminal amino acid of the mature protein, LMW-GS were classified into three types: LMW-i, LMW-m and LMW-s, corresponding to isoleucine, methionine and serine, respectively (D'Ovidio and Masci 2004). Each of the *Glu-A3* and *Glu-B3* loci encoded more than one type of LMW-GS and

was inferred from studies in which LMW-m-type genes were cloned from both *Glu-A3* (Lee et al. 1999; Ikeda et al. 2002) and *Glu-B3* loci (Huang and Cloutier 2008), respectively. The gene copy numbers of LMW-GS were estimated to vary from 10–15 copies (Harberd et al. 1985) or 35–40 copies (Cassidy et al. 1998; Sabelli and Shewry 1991) in hexaploid wheat. Furthermore, two LMW-GS genes could be separated from each other by more than 150 kb apart (Wicker et al. 2003), with the average distance estimated to be 81 kb (Huang and Cloutier 2008). More recently, Huang and Cloutier (2008) found nine LMW-m-type genes among the 12 LMW-GS genes identified from hexaploid wheat 'Glenlea' BAC library. Further, by employing nulli-tetrasomic lines of Chinese spring reported that seven of the nine LMW-m-type genes were located in the D genome.

It is important to explore the expression profiles of LMW-GS genes because of their influence on bread-making quality. However, illustrative reports so far are still limited. Several reports showed different expression models of LMW-GS genes, which could have resulted from the sensitivity of the detecting methods and errors of experiments (Grimwade et al. 1996; Altenbach 1998; Panozzo et al. 2001). In the past several years, quantitative real-time PCR (qRT-PCR) has been well established with apparent technical advantages and provided true quantitation of gene expression (Gachon et al. 2004). Recent studies reported on the transcription profiles of hordein and gliadin genes by qRT-PCR (Pistón et al. 2004, 2005, 2006); however, similarly detected profiles of LMW-GS genes in developing wheat grain endosperm are still lacking.

Aegilops species, closely related to bread wheat, possess unique diploid genomes classified as S^s, C, N, U, M and D^t. These genomes showed extensive allelic variations in storage compositions (Yan et al. 2003; Jiang et al. 2008), which could provide useful information for wheat quality improvement and evolutionary relationships. In the present study, two subclasses of LMW-m-type genes were isolated and characterized from different *Aegilops* species, and the bread wheat cultivar Zhongyou 9507 and their comparative transcriptional profiles during different grain developmental stages were monitored with qRT-PCR. The results obtained are expected to be beneficial to further understand the genomic organization of LMW-GS genes at the *Glu-3* loci and their expression characteristics in grain development.

Materials and methods

Plant material

Ae. uniaristata PI554419 ($2n = 2x = 14$, NN), *Ae. umbellulata* PI222762 ($2n = 14$, UU), *Ae. markgrafii* PI254863 ($2n = 14$, CC), *Ae. comosa* PI551017 ($2n = 14$, MM) and *Ae. speltoides* PI170204 ($2n = 14$, S^sS^s) were collected from CIMMYT, Mexico. The winter bread wheat cultivar Zhongyou 9507 with superior gluten properties (He et al. 2004) was kindly provided by the Institute of Crop Science, Chinese Academy of Agricultural Sciences of China.

Cloning of LMW-GS genes

Extraction of gDNA and amplification from dry seeds of aegilops accessions

Preparation method of gDNA for PCR amplification was based on An et al. (2006). The primers used were as follows:

Primer 1: 5′-ATC ATC ACA AGC ACA AGC ATC-3′;
Primer 2: 5′-TTC TTA TCA GTA GGC ACC AAC-3′;
Primer 3: 5′-ATG AAG ACC TTC CTC GTC TTT G-3′;
Primer 4: 5′-TCA GTA GGC ACC AAC TCC GGT AC-3′;
Primer 5: 5′-TCA GTA GGC ACC AAC TCC GGT GC-3′.

The amplification from Primer 1 + 2 included the upstream sequence of LMW-GS gene of about 50 bp. This primer pair amplifies LMW-GS genes from both *Aegilops* and *Triticum*, whereas primer combinations 3 + 4 and 3 + 5 were prone to amplifying the coding sequence of LMW-GS genes of *Triticum*.

Extraction of mRNA from premature seeds

Zhongyou 9507 was first vernalized and grown in the greenhouse at a temperature range of 20–25°C and later harvested following the method described by Altenbach (1998). Seeds from the central part of spikes were harvested at 5, 11, 14, 17, 20, 23, 26 and 29 DPA. Whole grains were ground in liquid N_2, homogenized in 0.1 ml volume, followed by resuspension of the tissue in 1 ml Trizol reagent after washing and was left to stand at 15°C for 10 min. After addition of 0.2 ml chloroform, the tube was shaken vigorously by hand for 15 s and incubated at 15°C for 3 min. After centrifugation at 4°C and 10,000*g* for 10 min, the aqueous phase was transferred exclusively to a clean tube. Isopropyl alcohol (0.4 ml) was further added to precipitate the RNA, after which the tube was kept for more than 1 h at −20°C. After centrifugation at 4°C and 10,000*g* for 10 min, the pellet was washed twice with 1 ml of 75% ethanol. Finally, the resultant RNA pellet was dried briefly and dissolved in RNase-free water.

Purification of mRNA

Contaminating DNA was eliminated by deoxyribonuclease (DNase). A total of 30 μl RNA (20–50 μg) was incubated with 15 U DNase I (TaKaRa, Dalian, Japan) at 37°C for 30 min, and 20 U inhibitor (TaKaRa) was simultaneously used to inactivate ribonuclease (RNase) activity. The digested products were mixed with 50 μl of ddH_2O treated with diethylpyrocarbonate (DEPC) and 100 μl of reagent I (phenol:chloroform:isoamyl alcohol in a volumetric ratio of 25:24:1). The mixture was first shaken vigorously for 10 min, followed by centrifugation at 4°C and 12,000*g* for another 10 min. The supernatant was transferred into a clean tube and 100 μl of reagent II (chloroform:isoamyl alcohol in a volumetric ratio of 24:1) was added into the supernatant followed by vigorous shaking for 10 min. After centrifugation at 4°C and 12,000*g* for 10 min, the supernatant was transferred into a clean column and 0.1 volume NaAc (3 M, pH 5.2) was subsequently added. Cooled ethanol of twice the total volume was added after which the mixture was incubated at −20°C for 1 h. After centrifugation at 4°C and 12,000*g* for 10 min, the pellet was washed with 70% ethanol, dried briefly and dissolved in RNase-free water. The RNA sample was run on a denaturing agarose gel to verify its integrity.

Synthesis of cDNA and PCR amplification

High quality RNA was used to synthesize cDNA with OligdT primer from approximately 100 ng mRNA by using a superscript first-strand synthesis (Promega, Madison, WI, USA).

Molecular cloning and sequencing

PCR products were separated on 1.0% agarose gels and the expected fragments were purified from the gels by using a Gel extraction kit (Omega, Doraville, GA, USA). Subsequently, the amplified nucleotides were ligated into pGEM-T Easy vector (Promega) and transformed into cells of *Escherichia coli* strain DH5α. By means of identification of bacterial PCR and digestion of restriction endonuclease, three positive clones were selected and sequenced on an automatic sequencer (TaKaRa).

Sequence alignment and phylogenetic analysis

The nucleotide sequences of cloned genes and deduced amino acid sequences were aligned by using the BioEdit 7.0 software. The phylogenetic tree and neighbor-joining tree were constructed by using the genes coding domain of the LMW-GS genes by DNAMAN 5.2.2 and SMART 4 softwares (Schultz et al. 1998).

Expression in *E. coli* and detection of fusion proteins

The expression vector was constructed with the plasmid pET-30a. The expression primer sequences were as follows:

Exp1: 5′-AAGCCATGGCTATGGAGACTAGATG-3′ (*Nco*I);
Exp2: 5′-AAGGAATTCTCAGTAGGCACCAAC-3′ (*Eco*RI);
Exp3: 5′-ATTCCATGGCTATGAAGACCTTCC-3′ (*Nco*I).

To facilitate identification of expressed proteins by Western blotting, the His tag and S tag lying behind the T7 promoter were reserved together and the fusion tag was 5,038 Da. Transforming and the inducing procedure of heterologous expression, as well as the identification of fusion proteins with SDS-PAGE, Western blotting and MALDI-TOF-MS, were performed according to Li et al. (2007) and An et al. (2006).

Quantitative real-time PCR

A tenfold serial dilution series of plasmids containing sequences of two typical genes from two LMW-m subclasses, respectively, were used as standard samples to construct standard curves. The mass concentrations of the plasmid were measured by NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) at 260 nm (A260) wavelength and then converted to the copy concentration by using the following equation (Whelan et al. 2003):

$$\text{DNA (copy)} = \frac{6.02 \times 10^{23}\ (\text{copies mol}^{-1}) \times \text{DNA amount (g)}}{\text{DNA length (bp)} \times 660\ (\text{g mol}^{-1}\text{bp}^{-1})}.$$

The amount of total RNA purified was also quantified with the NanoDrop ND-1000 spectrophotometer in the A260/A280 and A260/A230 ratio.

Quantitative real-time PCR was performed on Rotor gene 3000 (Corbett Research, Sydney, Australia). The reaction was carried out in 20 μl with the following final concentrations: SYBR green Realmastermix (Tiangen, Beijing, China), 1×, 0.2 μM of each primer and 2 μl cDNA. The PCR conditions were as follows: 94°C for 3 min, 40 cycles of 94°C for 20 s, 59°C for 15 s and 68°C for 20 s. After the end of the last cycle, while maintaining the temperature at 68°C for 1 min, the products were heat denatured over a temperature gradient at 0.1°C s^{-1} from 72 to 95°C. The fluorescence signal was collected at the end of extension in every cycle. The standard curve method of

Rotor-Gene analysis 6.1 software was used to analyze the quantity of target gene.

Results

Cloning and characterization of two subclasses among LMW-m-type genes

PCR products from the different species are shown in Fig. 1. A single band with about 900 bp was amplified by different allele-specific primers that correspond to the size of LMW-GS genes. Six complete genes from *Aegilops* species (two from M genome and one from each of C, N, S and U genomes) and that from wheat cultivar Zhongyou 9507 were sequenced. From the 12 LMW-GS genes that were sequenced, 7 with greater sequence variations and concurrently possessing the typical LMW-GS structural characters were deposited in GenBank and subsequently used for sequence comparative analysis (Table 1). Their deduced amino acid sequences showed that all belonged to LMW-m-type subunits due to M (methionine) as the first residue in the mature protein.

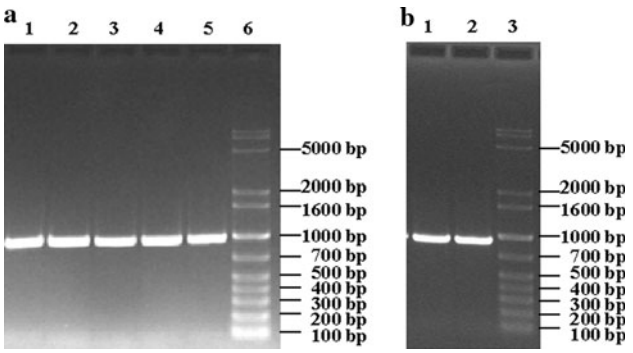

Fig. 1 PCR amplification of LMW-GS genes from *Aegilops* and *Triticum* species. **a** *Lanes 1*, *2*, *3*, *4* and *5* PCR products amplified from *Ae. comosa* (PI551017), *Ae. umbellulata* (PI222762), *Ae. speltoides* (PI170204), *Ae. markgrafii* (PI254863) and *Ae. uniaristata* (PI554419) by the Primer 1 + 2, respectively. *Lane 6* the 1 kb plus marker. **b** *Lane 1 and 2* PCR products amplified from *T. aestivum* (Zhongyou 9507) by the Primer 3 + 4 and Primer 3 + 5, respectively. *Lane 3* the 1 kb plus marker

The deduced amino acid sequences of seven novel genes were aligned with ten other LMW-m types previously isolated (Fig. 2). As typical for LMW-GS genes, all contained signal peptide of 20 amino acid residues, N-terminal domain, repetitive domain and conservative domain as well as eight conservative cysteine residues. As shown in Fig. 2, 17 LMW-m-type genes could be clearly classified into two subclasses, designated as 1 and 2, with each subclass showing high similarity within themselves. For instance, *ZyLMW-m1* differed by only three amino acid residue substitutions compared to the others in the same subclass. However, several discrepancies were evidenced between the two subclasses, including single-nucleotide polymorphisms (SNPs) and insertions/deletions (InDels) variations. Gene structural comparison and distinguishing characteristics between the two subclasses are shown in Table 2. Compared with subclass 1, subclass 2 had 21 amino acid residue substitutions, two fragment deletions (each with seven amino acid residues) and a double-residue deletion as well as two fragment insertions (12 and 2–5 residues). Furthermore, the seventh cysteine residue was located at the second position of the 12 residue insertion fragment in subclass 2, whereas the positions of the other seven cysteine residues were highly conserved. Of the 12 LMW-GS genes isolated in this study, 2 and 10 genes were clustered into subclass 1 and 2, respectively. The gene from the N genome was classified into subclass 1, while those from C, M, S and U genomes belonged to subclass 2. This result confirmed that the related genomes characterized exhibited the *Glu-3* loci as those found in the A, B and D genomes of hexaploid bread wheat.

Phylogenetics and genome evolution among bread wheat and *Aegilops* species

To further investigate the phylogenetic relationships among LMW-GS genes from different genomes, 7 cloned LMW-GS genes as well as other 35 genes deposited in GenBank, in which their repetitive domains were removed,

Table 1 Seven LMW-GS genes isolated from *Aegilops* and *Triticum* species

LMW-GS genes	GenBank Accession No.	Size (bp)	Deduced subunit *Mr* (Da)	Genome	Species
AuLMW-m1	EU571724	936	33,357.6	N	*Ae. uniaristata* PI554419
AumLMW-m1	EU571725	891	31,345.22	U	*Ae. umbellulata* PI222762
AmLMW-m1	EU329425	900	31,724.6	C	*Ae. markgrafii* PI254863
AcLMW-m1	EU594338	891	31,188.9	M	*Ae. comosa* PI551017
AsLMW-m1	EU571721	897	31,665.6	S^s	*Ae. speltoides* PI170204
ZyLMW-m1	EU329426	915	32,593.8	D	*T. aestivum* Zhongyou 9507
ZyLMW-m2	EU329427	912	32,045.8	D	*T. aestivum* Zhongyou 9507

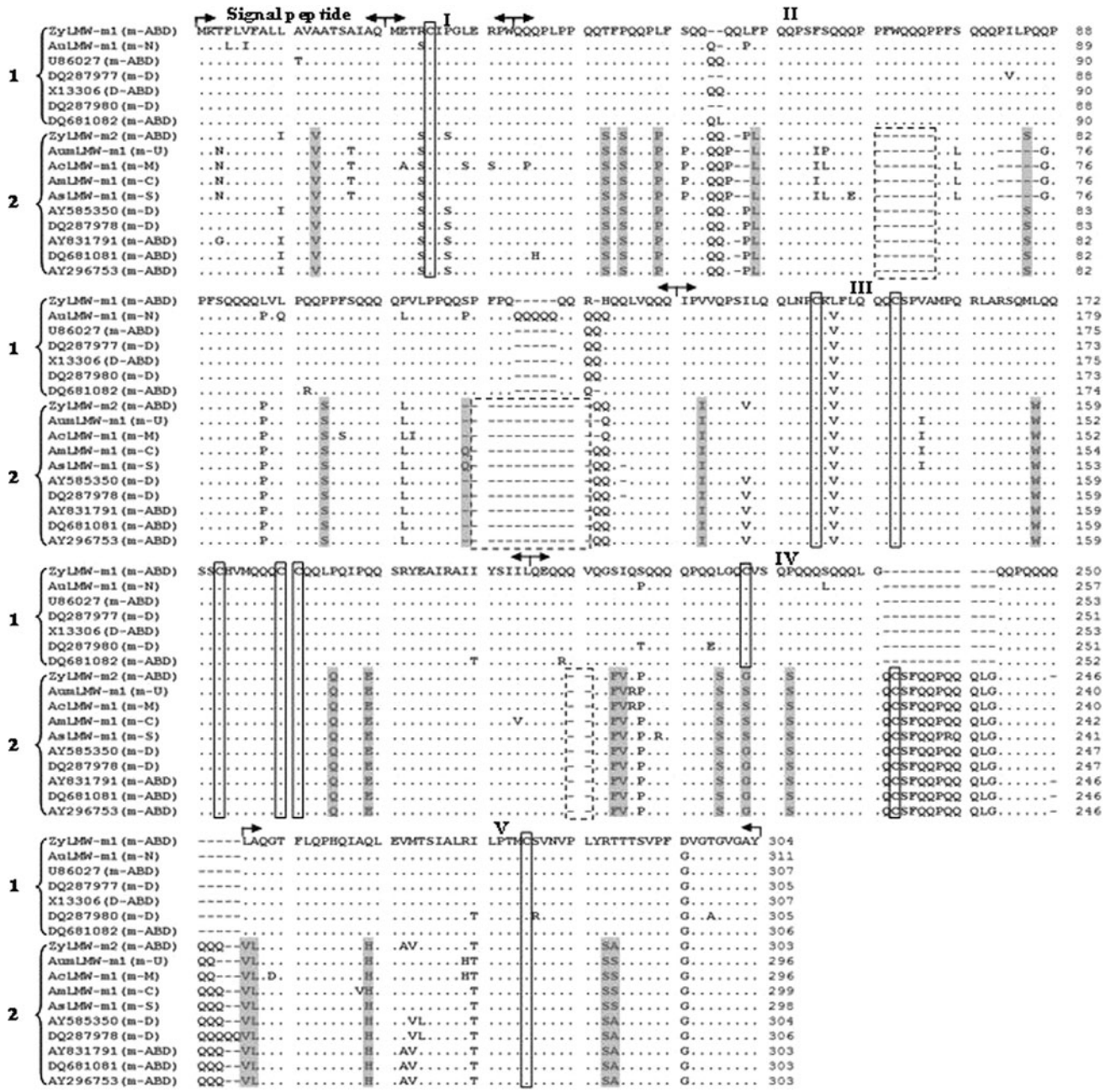

Fig. 2 Comparison of deduced amino acid sequences between seven novel genes isolated and the other ten LMW-m-type genes obtained from GenBank. The mature protein sequences of LMW-GS are divided into five domains: *I* N-terminal domain; *II* repetitive domain. *III* cysteine-rich region. *IV* glutamine-rich region. *V* C-terminal conservative region. The same sequences and deletions with ZyLMW-m1 subunit are shown by dots and dashes, respectively. The cysteine residues are shown with *black frames*. The *shadows* indicate a single amino acid substitution and a double amino acid inversion between the two subclasses. The *broken black frames* show deletions. The GenBank Accession numbers are as follows: AY296753 (Long et al. 2004), AY585350 (Johal et al. 2004), AY831791 (Ozdemir and Cloutier, 2005), DQ287977 (Pei et al. 2007), DQ287978 (Pei et al. 2007), DQ287980, DQ681081, DQ681082, U86027 and X13306 (Colot et al. 1987). The LMW-GS gene types and their genomes are shown in *bracket*, and *ABD* indicates the gene locations not determined

were used to construct a phylogenetic tree (Fig. 3). It was demonstrated that all 42 genes could be clustered into two branches, well corresponding to LMW-m-type and LMW-i-type genes, respectively. Apparently, the LMW-i-type genes showed the most variations among LMW-GS genes, with 77% similarity to LMW-m and LMW-s-type genes. In particular, the LMW-m-type genes could be clearly separated into two subclasses (subbranch 2a and 2b in Fig. 3)

Table 2 Structural features of two subclasses of LMW-m subunits

Subclass	13	43	45	49	59	71–77	87	104	119	120–131	143	178	195	199
1	A	T	P	L	F	PFWQQQP	P	P	S/P	PFPQ(QQQQQ)QQQ/R	V	L	P	Q
2	V	S	S	P	L	–	S/–	S	Q/–	–	I	W	Q	E

Subclass	220–221	224–225	235	238	242	252–263	271–275	276–277	289	313–314
1	QV	SI	L	C	P	–	–	LA	Q	RT
2	–	FV	S	G/S	S	QCSFQQPQQQLG	QQ(QQQ)	VL	H	SA/S

with 90% similarity, which was well consistent with the results of sequence comparative analysis as described above. Eight LMW-s-type genes were also classified into subclass 1, suggesting that they were closely related to LMW-m-type genes. Actually, both LMW-s and LMW-m-type genes had similar N-terminal sequence and high homology. The N-terminal sequence (MEN)SHIPG and (IEN)SHIPG of LMW-s subunits might result from post-translational processing, and therefore they could be derivatives of LMW-m-type subunits (Masci et al. 1998; Huang and Cloutier 2008).

As shown in Fig. 3, the *ZyLMW-m1* and *ZyLMW-m2* genes from bread wheat Zhongyou 9507 were separately classified into subclass 1 and 2, respectively, and in addition could be assigned to the D genome due to their high similarity to genes from *Ae. tauschii*. These suggest that the LMW-m-type genes encoded by *Glu-D3* locus in bread wheat diverged into two subclasses during the evolutionary process. The estimated gene divergent times between seven LMW-GS genes encoded by different genomes are listed in Table 3. The *ZyLMW-m1* and *ZyLMW-m2* genes encoded by D genome diverged early, occurring at 6.85 (±0.85) MYA. The divergence between *ZyLMW-m2* and those from C, M, S^s and U genomes, as well as *ZyLMW-m1* and the gene from N genome, occurred at about 2–3.5 MYA. As revealed by estimated gene divergent times, the C, M, N and U genomes were more related to D genome and their divergence occurred more recently. It could be deduced that the divergent time between two LMW-m gene subclasses was much earlier than those between genomes' divergence.

Expression of the LMW-GS genes in *E. coli*

The *ZyLMW-m1* and *ZyLMW-m2* genes possessing typical characteristics of LMW-m subclass 1 and 2, respectively, were used for heterologous expression in *E. coli*. The sequences coding mature proteins of both genes were amplified with primers Exp1 + Exp2 and Exp1 + Exp3, respectively, and ligated into the pET-30a expression vector. The fusion proteins were identified by both SDS-PAGE and Western blotting. As shown in Fig. 4a, the induced fusion proteins of both *ZyLMW-m1* and *ZyLMW-m2* genes could be identified by SDS-PAGE (indicated by arrows). Western blotting showed that the induced proteins extracted from *E. coli* strongly hybridized to the anti-His Tag mouse monoclonal antibody, but without any signal to bacterium (Fig. 4b). The expression proteins were further detected by MALDI-TOF-MS, and two single subunits with 36,139.6 and 35,579.9 Da (including 5,038 Da fusion tag), corresponding to ZyLMW-m1 and ZyLMW-m2 subunits, respectively, were identified (Fig. 4c, d). This confirmed that the hybridized proteins could be the expressed products of the two fusion genes, and hence both genes were able to express successfully in *E. coli*. However, the deduced molecular weights (32,593.8 and 32,045.8 Da) of *ZyLMW-m1* and *ZyLMW-m2* genes were 944.2 and 1,492.2 Da higher than those of their expression proteins, respectively. A similar result was reported previously (Lee et al. 1999), in which the expression of proteins in *E. coli* moved faster than the native subunits on SDS-PAGE gels. In the present study the structural differences between the expressed proteins in *E. coli* and native proteins in the seed endosperm could result in their molecular weight changes.

Transcriptional analysis of LMW-GS genes in different development periods of seed endosperm in Zhongyou 9507

Quantitative real-time PCR (qRT-PCR) was a highly sensitive technique and suitable to detect the mRNA amount of storage protein genes expressed in the developmental endosperm (Pistón et al. 2004, 2005, 2006). In the present study, the transcription levels of two novel LMW-m genes, *ZyLMW-m1* and *ZyLMW-m2*, from subclass 1 and 2, respectively, were monitored by qRT-PCR. Two pairs of primers were selected to amplify the special fragment sequences of two genes as shown in Table 4. The specificity of each pair of primers was tested with melting curve and agarose gel electrophoresis. Unique melting temperature peak and single product bands (146 bp) from two LMW-GS genes are shown in Fig. 5, which indicated that non-specific amplifications were obtained.

Both curves were highly linear ($R^2 > 0.999$) in the range tested by duplicate reactions. The slopes of the standard

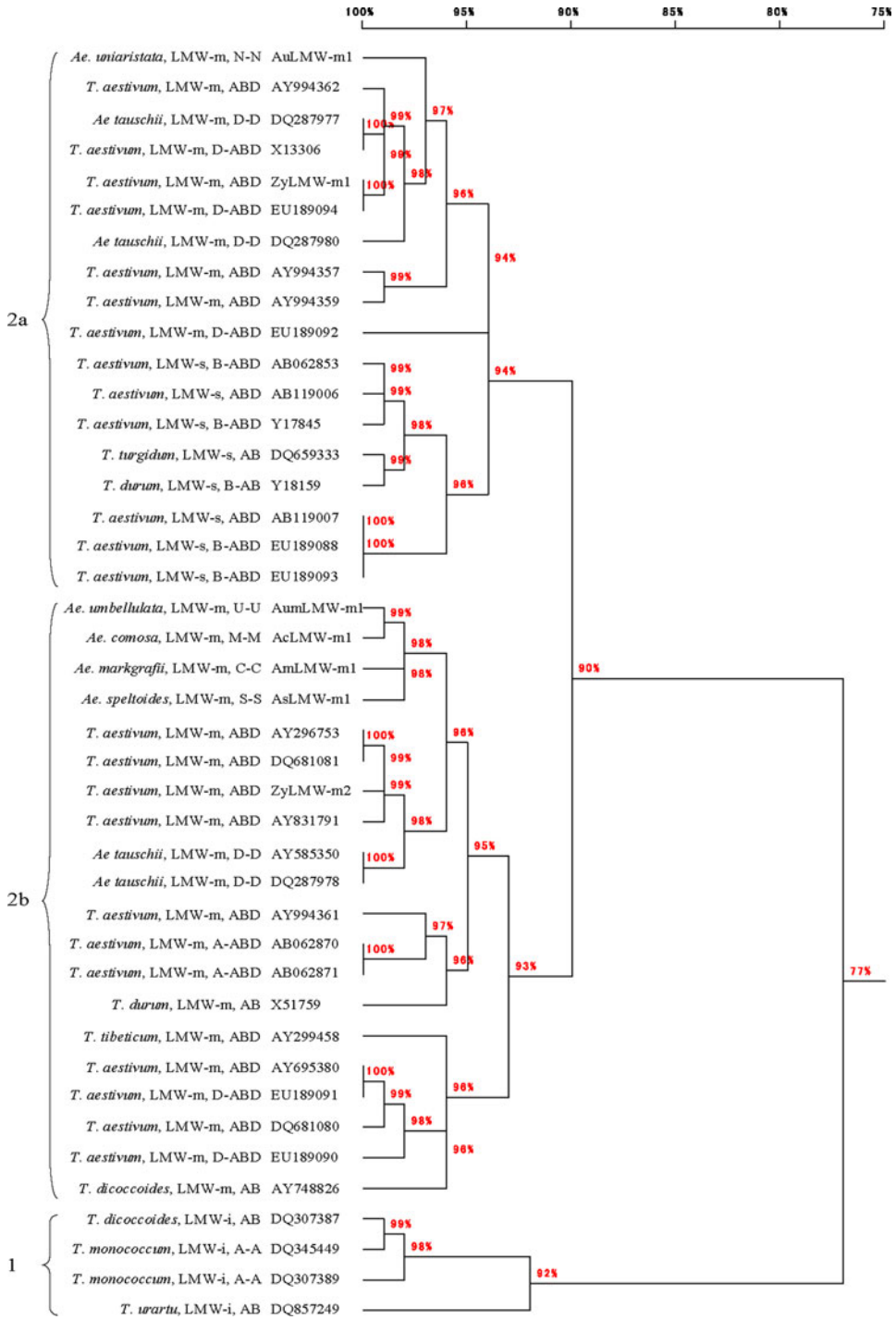

Fig. 3 Phylogenetic tree constructed with 42 LMW-GS genes, in which the repetitive domain sequences were removed. The GenBank Accession numbers are as follows: AB062853, AB062870, AB062871 (Ikeda et al. 2002), AB119006, AB119007 (Maruyama-Funatsuki et al. 2005), Y17845 (Masci et al. 1998), DQ659333, Y18159 (D'Ovidio et al. 1999), EU189088, EU189090, EU189091, EU189092, EU189093, EU189094 (Huang and Cloutier, 2008), AY994357, AY994359, AY994362, DQ287977, DQ287978, X13306, DQ287980, AY296753, DQ681081, AY831791, AY585350, AY994361 (Zhao et al. 2004), X51759, AY299458(Wang et al. 2005), AY695380, DQ681080, AY748826, DQ307387 (LMW-i), DQ345449 (LMW-i), DQ307389 (LMW-i, An et al. 2006), DQ857249 (LMW-i) and the seven new genes from this study as shown in Fig. 2. The LMW-GS gene types, species and their genomes are indicated

Table 3 Estimation of divergence time among seven LMW-GS genes (MYA)

LMW-GS genes	1	2	3	4	5	6
1. *AuLMW-m1*(N)						
2. *AumLMW-m1*(U)	6.69 ± 0.85					
3. *AmLMW-m1*(C)	6.85 ± 0.85	1.15 ± 0.38				
4. *AsLMW-m1*(S^s)	7.08 ± 0.92	1.38 ± 0.38	1.23 ± 0.38			
5. *AcLMW-m1*(M)	7.31 ± 0.85	0.62 ± 0.23	1.77 ± 0.46	2.00 ± 0.46		
6. *ZyLMW-m1*(D)	2.00 ± 0.46	6.31 ± 0.77	6.54 ± 0.85	6.54 ± 0.85	6.92 ± 0.85	
7. *ZyLMW-m2*(D)	6.85 ± 0.85	2.92 ± 0.54	3.00 ± 0.54	3.31 ± 0.54	3.54 ± 0.62	6.85 ± 0.85

The repetitive domains of gene-coding regions were removed. Bootstrap percentages were based on 1,000 iterations

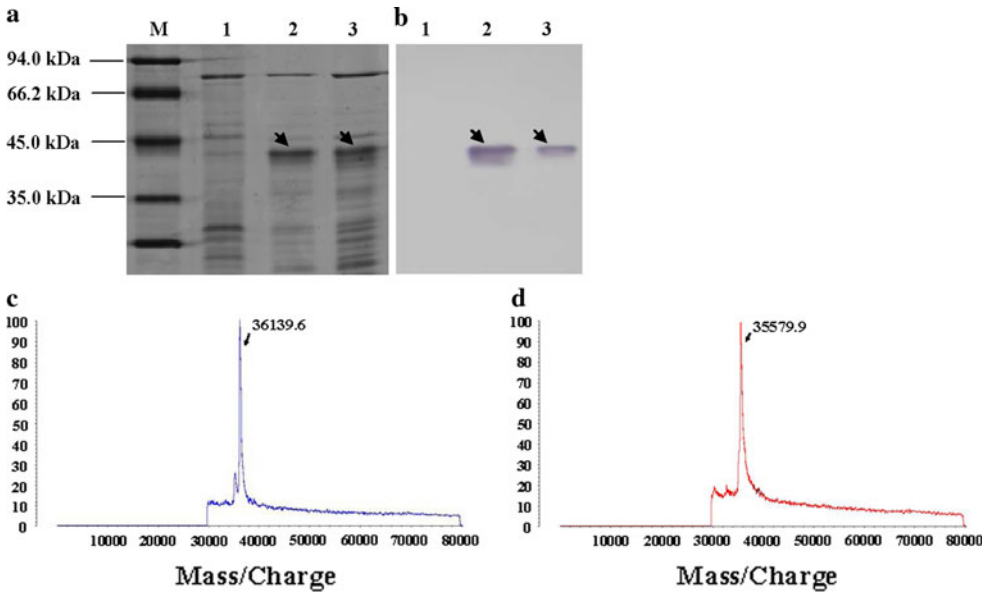

Fig. 4 Induced products in *E. coli* identified and detected by SDS-PAGE (**a**), Western blotting (**b**) and MALDI-TOF-MS (**c**, **d**). **a** *M* the protein marker; *lane 1* empty vector pET-30a; *lane 2* ZyLMW-m1; *lane 3*, ZyLMW-m2. **b** *Lane 1* empty vector pET-30a; *lane 2* ZyLMW-m1; *lane 3* ZyLMW-m2. **c** ZyLMW-m1. **d** ZyLMW-m2

Table 4 Specific primers designed for qRT-PCR

LMW-GS genes	Primer	Sequence
ZyLMW-m1	$R_{m1}1$	5′-TGCAGCCACACCAGATAGCTCAG-3′
	$R_{m1}2$	5′-TCAGTAGGCACCAACTCCGGTAC-3′
ZyLMW-m2	$R_{m2}3$	5′-TGCAGCCACACCAGATAGCTCAC-3′
	$R_{m2}4$	5′-TCAGTAGGCACCAACTCCGGTG-3′

curves for *ZyLMW-m1* and *ZyLMW-m2* were −3.21 and −3.38, respectively. From the slopes, amplification efficiencies of 1.05 and 1.02 were determined for *ZyLMW-m1* and *ZyLMW-m2* in the investigated range, respectively.

Although the quantity of mRNA was very low, the mRNA of the two genes could be detected at the first test period (5 DPA) as shown in Fig. 6. This indicated that the transcription of LMW-GS genes was not later than 5 DPA in Zhongyou 9507. Both *ZyLMw-m1* and *ZyLMW-m2* reached the maximum transcription level at 14 DPA and then decreased from 14 to 29 DPA. In the following three periods at 17, 20 and 23 DPAs, the mRNA was stably transcribed at low levels. The mRNA quantity of *ZyLMW-m2* from 11 to 14 DPA was more than ten times higher than those of *ZyLMw-m1*, suggesting that the transcription level in two LMW-m subclasses was probably different. In addition, it was noted that the transcription quantity of *ZyLMW-m2* gene was much higher than that of the *ZyLMW-m1* gene in Zhongyou 9507 in all the grain developing stages analyzed.

Discussion

Genomic organization and evolution at *Glu-3* loci among bread wheat and *Aegilops* species

LMW-GS genes belong to a multiple gene family and possess multiple copies in *Triticum aestivum* and related wild species. It was estimated that the number in hexaploid bread wheat varied from 10–15 copies (Harberd et al. 1985), to 35–40 copies (Sabelli and Shewry 1991; Cassidy et al. 1998), and most LMW-GS genes were located on the short arms of the group 1 chromosomes (D'Ovidio and

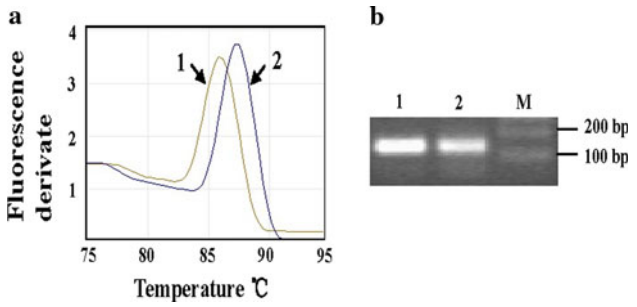

Fig. 5 Identification of specific primers for qRT-PCR. Melting curves (**a**) and agarose gel electrophoresis (**b**) of the amplification products from qRT-PCR using the specific primer pairs for two novel LMW-GS genes. *Lanes 1 and 2* are the amplifications of the *ZyLMW-m1* and *ZyLMW-m2* genes, respectively

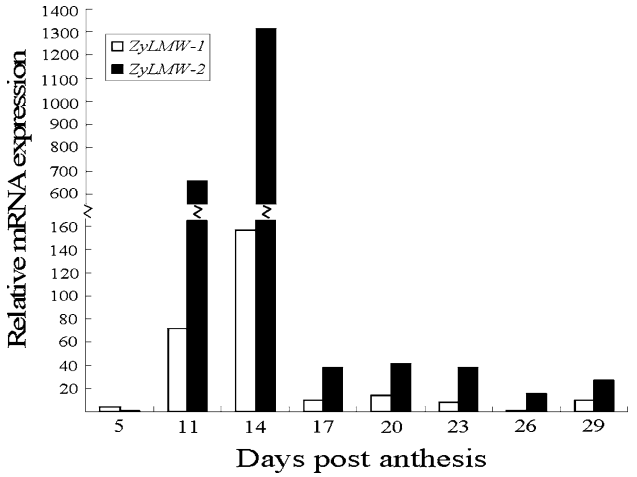

Fig. 6 The expression profiles and relative transcript quantity of two novel LMW-GS genes *ZyLMW-m1* and *ZyLMW-m2* during five different development stages of endosperm detected by qRT-PCR

Masci 2004). Although the exact copies are unknown so far, it is apparent that each *Glu-3* locus encodes multiple LMW-GS genes. More recently, Huang and Cloutier (2008) identified 12 LMW-GS genes from the hexaploid wheat 'Glenlea' BAC library and found that most of them were located in the D genome.

In the present study, 12 LMW-m genes were isolated from the Chinese bread wheat cultivar Zhongyou 9507 and five Aegilops species, among which their structural features revealed that the related C, M, S, N and U genomes had a similar *Glu-3* locus as in A, B and D genomes of hexaploid wheat. Comparative analysis analyzed in combination with previously characterized LMW-GS genes in bread wheat and related species demonstrated that two subclasses (1 and 2) were in existence, with each possessing significant difference in structural characteristics (Table 2, Fig. 2). Among six LMW-m genes characterized from Zhongyou 9507, which were assigned to the D genome, one and five genes were classified into subclass 1 and 2, respectively. The divergence between these two subclasses in bread wheat occurred at about 6 MYA, generally earlier than the divergent times of the five related *Aegilops* genomes as shown in Table 3 and Fig. 3. This suggested that the divergence of the two subclasses of LMW-m-type gene at *Glu-D3* locus could be earlier than those between D and C, M, S or U genomes.

To date, the origin of C, M, S, N and U genomes from *Aegilops* species and their evolutionary relationships with A, B and D genomes of common wheat are not clarified. Our recent investigation (unpublished data) indicated that the C, N and U genomes had close lineage and probably been derived from the same progenitor. It is widely accepted that the A genome of common wheat originated from *Triticum urartu* and the D genome from *Ae. tauschii* (Petersen et al. 2006), whereas the origin of the B genome remains controversial (Huang et al. 2002). Considerable studies support that the B genome originated from S^s genome of *Ae. speltoides* (Fernandez-Calvin and Orellana 1990; Liu et al. 1999; Petersen et al. 2006), but studies on genes encoding plasmid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex questioned again the origin of the B genome (Huang et al. 2002). Our results from LMW-m genes encoded by *Glu-3* loci in *Triticum* and *Aegilops* species showed close genetic relationship between B and S^s genomes and therefore supported the conception that the B genome may have been derived from S^s genome. The C, M, N and U genomes appeared to be more related to the D genome of bread wheat and their divergence occurred at about 2–6 MYA (Table 3). In addition, since two subclasses of LMW-m genes were isolated synchronously from D genome of bread wheat, we speculate that two LMW-m gene subclasses may also be present in the C, M, N and U genomes due to their close phylogenetic relationships to the D genome.

Our present finding is also complemented by the results of Huang and Cloutier (2008). In their study, four LMW-m-type genes assigned to the D genome (GenBank Accession No. EU189090, EU189091, EU189092 and EU189094) were identified from bread cultivar 'Glenlea' through BAC library sequencing. The same GenBank accessions, which we also used in the present study, showed that the former two genes clustered into subclass 1 and the latter two in subclass 2, respectively (Fig. 3). The *Glu-D3* locus in the D genome of bread wheat might encode the most LMW-GS genes, and therefore could be the largest locus among three *Glu-3* loci. The traditional method classifies LMW-GS genes into three types: LMW-s, LMW-m and LMW-i based on the first amino acid residue of the N terminus (D'Ovidio and Masci 2004). So far, the A, B and D genomes of *T. aestivum* as well as their related genomes in the wild species have been shown to encode two kinds of LMW-GS genes. For instance, the *Glu-A3* encodes not only multiple LMW-i-type genes

(Wicker et al. 2003; An et al. 2006), but also the LMW-m-type genes (Lee et al. 1999; Ikeda et al. 2002). Similarly, both LMW-s and LMW-m-type genes were identified from *Glu-B3* (Huang and Cloutier 2008) and *Glu-D3* (Zhao et al., 2006). Furthermore, our present results showed that the *Glu-D3* locus in bread wheat encoded two subclasses of LMW-m-type genes, suggesting that the *Glu-D3* locus might have more extensive allelic variations and hence might encode the most LMW-GS genes among the three *Glu-3* loci in bread wheat, thus complementing a postulation by Huang and Cloutier (2008) that the *Glu-D3* locus is larger than the *Glu-B3* and *Glu-A3* loci. Previously, the LMW-s-type gene was considered to be predominant in *T. aestivum* (Lew et al. 1992), but most of recently isolated LMW-GS genes from *T. aestivum* and *Ae. tauschii* (Pei et al. 2007; Huang and Cloutier 2008) and other wheat relatives characterized in this study belonged to LMW-m-type genes. Furthermore, most of the LMW-GS genes deposited in GenBank so far belonged to subclass 2 according to our analysis (results not shown). Therefore, we speculate that the subclass 2 LMW-m gene could be the most common type among *Triticum* and related wild *Aegilops* species, and both LMW-s and LMW-i genes could be the variant types of LMW-m genes. It is most likely that the LMW-s-type gene had diverged from the subclass 1 LMW-m-type as evidenced from their close relationships shown in Fig. 3.

In terms of the molecular mechanisms of the origin and evolution of storage protein genes, it is known that the extensive allelic variations are mainly from SNPs and InDels, generally resulting from unequal crossing over, slip-mismatching and point mutations (Anderson and Greene 1989; An et al. 2006; Zhang et al. 2006). More recently, it has been found that, in addition to unequal homologous recombination, illegitimate recombination appears to be an important genetic mechanism for the origin and evolution of *Glu-3* alleles as well as *Glu-1* and other prolamine genes (Zhang et al. 2008; Li et al. 2008). Since two or different types of LMW-GS genes were found to locate separately at a single *Glu-3* locus, these could facilitate the occurrence of illegitimate recombination events (Wicker et al. 2003; Li et al. 2008), which probably result in the generation of new alleles and rapid divergence of storage protein loci and genomes in *Aegilops* and *Triticum* species.

Transcription and protein synthesis of LMW-GS genes during grain development

Until now, several different results have been reported on the expression profiles of LMW-GS genes in the developmental endosperm. It appears that the sensitivity of detecting methods for mRNA and proteins and genotype differences may have been mainly responsible for the inconsistency of the LMW-GS expression model. Panozzo et al. (2001) showed that the LMW-GS deposition was rapidly increased after 14 DPA in endosperm and its synthesis was later than that of HMW-GS. Other investigations revealed that LMW-GS began its synthesis in endosperm from 5 to 13 DPA (Ng et al. 1991; Gupta et al. 1996; Grimwade et al. 1996; Kawaura et al. 2005). Altenbach (1998) also detected transcript products of LMW-GS genes in 15 DPA by competitive RT-PCR. In this study, small mRNA amounts of both *ZyLMW-m1* and *ZyLMW-m2* genes were detected at 5 DPA by qRT-PCR, displaying their early expression in comparison with the expression at 8 DPA for the γ- and B-hordeins from *Hordeum chilense* as well as γ-gliadins from bread wheat (Pistón et al. 2004, 2005, 2006). Both of the LMW-m genes reached the maximum level in transcription amount at 14 DPA, later than that of γ3-hordein gene at 12 DPA (Pistón et al. 2004), but earlier than that of γ-gliadin genes of I, III and IV groups at 18 DPA (Pistón et al. 2006). Kawaura et al. (2005) reported that LMW-GS genes encoded by D genome reach the maximum level at 10 DPA by comprehensive expressed sequence tags (ESTs) of common wheat through four grain developmental stages. Interestingly, the mRNA quantity of *ZyLMW-m2* gene in 14 DPA was almost tenfold higher than that of the *ZyLMW-m1* gene (Fig. 6), suggesting a highly different transcription level among LMW-GS genes as well as a difference of expression between two gene subclasses. Pistón et al. (2008) found that the efficiency of promoter showed discrepancy caused by sequence among genes. Since the promoter sequences of LMW- and HMW-GS genes are highly conserved (Yan et al. 2004; An et al. 2006; Li et al. 2008), some specific structural features of LMW-GS genes could result in their expression differences. Particularly, some significantly different characters presented in subclass 2, such as two fragment insertions/deletions and the position change of the seventh cysteine residues should be responsible for the high level expression.

Acknowledgments This research was financially supported by grants from the National Natural Science Foundation of China (30830072, 30771334), the Chinese Ministry of Science and Technology (2009CB118303, 2006AA10Z186) and the Key Developmental Project of Science and Technology, Beijing Municipal Commission of Education (KZ200910028003).

References

Altenbach SB (1998) Quantification of individual low-molecular-weight glutenin subunit transcripts in developing wheat grains by competitive RT-PCR. Theor Appl Genet 97:413–421

An X, Zhang Q, Yan Y, Li Q, Zhang Y, Wang A, Pei Y, Tian J, Wang H, Hsam SLK, Zeller FJ (2006) Cloning and molecular characterization of three novel LMW-i glutenin subunit genes

from cultivated einkorn (*Triticum monococcum* L.). Theor Appl Genet 113:383–395

Anderson OD, Greene FC (1989) The characterization and comparative analysis of high-molecular-weight glutenin genes from genomes A and B of a hexaploid bread wheat. Theor Appl Genet 77:689–700

Cassidy BG, Dvorak J, Anderson OD (1998) The wheat low-molecular-weight glutenin genes: characterization of six new genes and progress in understanding gene family structure. Theor Appl Genet 96:743–750

Colot V, Robert LS, Kavanagh TA, Bevan MW, Thompson RD (1987) Localization of sequences in wheat endosperm protein genes which confer tissue-specific expression in tobacco. EMBO J 6:3559–3564

D'Ovidio R, Masci S (2004) The low-molecular-weight glutenin subunits of wheat gluten. J Cereal Sci 39:321–339

D'Ovidio R, Marchitelli C, Ercoli Cardelli L, Porceddu E (1999) Sequence similarity between allelic *Glu-B3* genes related to quality properties of durum wheat. Theor Appl Genet 98:455–461

Fernandez-Calvin B, Orellana (1990) High molecular weight glutenin subunit variation in the Sitopsis section of *Aegilops*. Implications for the origin of the B genome of wheat. Heredity 65:455–463

Gachon C, Mingam A, Charrier B (2004) Real-time PCR: what relevance to plant studies? J Exp Bot 55:1445–1454

Gianibelli MC, Larroque OR, MacRichie F, Wrigley CW (2001) Biochemical, genetic, and molecular characterization of wheat glutenin and its component subunits. Cereal Chem 78:635–646

Grimwade B, Tatham AS, Freedman RB, Shewry PR, Napier JA (1996) Comparison of the expression patterns of genes coding for wheat gluten proteins and proteins involved in the secretory pathway in developing caryopses of wheat. Plant Mol Biol 30:1067–1073

Gupta RM, Masci S, Lafiandra D, Bariana HS, MacRitchie F (1996) Accumulation of protein subunits and their polymers in developing grains of hexaploid wheats. J Exp Bot 47:1377–1385

Harberd NP, Bartels D, Thompson RD (1985) Analysis of the gliadin multigene loci in bread wheat using nullisomic-tetrasomic lines. Mol Gene Genet 198:234–242

He ZH, Yang J, Zhang Y, Quail KJ, Pena RJ (2004) Pan bread and dry white Chinese noodle quality in Chinese winter wheats. Euphytica 139:257–267

Huang XQ, Cloutier S (2008) Molecular characterization and genomic organization of low molecular weight glutenin subunit genes at the *Glu-3* loci in hexaploid wheat (*Triticum aestivum* L.). Theor Appl Genet 116:953–966

Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci USA 99:8133–8138

Ikeda TM, Nagamine T, Fukuoka H, Yano H (2002) Identification of new low-molecular-weight glutenin subunit genes in wheat. Theor Appl Genet 104:680–687

Jiang CX, Pei YH, Zhang YZ, Li XH, Yao DN, Yan YM, Hsam SLK, Zeller FJ (2008) Molecular cloning and characterization of four novel LMW glutenin subunit genes from *Aegilops longissima*, *Triticum dicoccoides* and *T. zhukovskyi*. Hereditas 145:92–98

Johal J, Gianibelli MC, Rahman S, Morell MK, Gale KR (2004) Characterization of low-molecular-weight glutenin genes in *Aegilops tauschii*. Theor Appl Genet 109:1028–1040

Kawaura K, Mochida K, Ogihara Y (2005) Expression profile of two storage-protein gene families in hexaploid wheat revealed by large-scale analysis of expressed sequence tags. Plant Physiol 139:1870–1880

Lee YK, Ciaffi M, Appels R, Morell MK (1999) The low-molecular-weight glutenin subunit proteins of primitive wheats. II. The genes from A-genome species. Theor Appl Genet 98:126–134

Lew EJL, Kuzmicky DD, Kasarda DD (1992) Characterization of low molecular weight glutenin subunits by reversed-phase high-performance liquid chromatography, sodium dodecyl sulfate-polyacrylamide gel electrophoresis, and N-terminal amino acid sequencing. Cereal Chem 69:508–515

Li XH, Zhang YZ, Gao LY, Wang AL, Ji KM, He ZH, Appel R, Ma WJ, Yan YM (2007) Molecular cloning, heterologous expression, and phylogenetic analysis of a novel y-type HMW glutenin subunit gene from the G genome of *Triticum timopheevi*. Genome 50:1130–1140

Li XH, Ma W, Gao LY, Zhang YZ, Wang AL, Ji KM, Wang K, Appels R, Yan Y (2008) A novel chimeric LMW-GS gene from the wild relatives of wheat *Ae. kotschyi* and *Ae. juvenalis*: evolution at the *Glu-3* loci. Genetics 180:93–101

Liu X, Wang RC, Jia J (1999) Genome relationship among Sitopsis species of *Aegilops* and the B/G genome of *Triticum* assessed by RAPD markers. In: Slinkard AE (ed) Proceedings of the 9th International Wheat Genetics Symposium. University Extension Press, Saskatoon, pp 79–81

Long H, Yan ZH, Wei YM, Zheng YL (2004) Molecular cloning of a novel low-molecular-weight glutenin subunit gene from wheat (*Triticum aestivum* L.), Variety 'Chuannong 16'. J Genet Genomics 30:1179–1184

Maruyama-Funatsuki W, Takata K, Funatsuki H, Tabiki T, Ito M, Nishio Z, Kato A, Saito K, Yahata E, Saruyama H, Yamauchi H (2005) Identification and characterization of a novel LMW-s glutenin gene of a Canadian Western extra-strong wheat. J Cereal Sci 41:47–57

Masci S, D'Ovidio R, Lafiandra D, Kasarda DD (1998) Characterization of a low-molecular-weight glutenin subunit gene from bread wheat and the corresponding protein that represents a major subunit of the glutenin polymer. Plant Physiol 118:1147–1158

Ng PKW, Slominski E, Johnson WJ, Bushuk W (1991) Changes in wheat endosperm proteins during grain maturation. In: Bushuk W, Tkachuk R (eds) Gluten proteins. American Association of Cereal Chemistry, St. Paul pp 740–754

Ozdemir N, Cloutier S (2005) Expression analysis and physical mapping of low-molecular-weight glutenin loci in hexaploid wheat (*Triticum aestivum* L.). Genome 48:401–410

Panozzo J, Eagles HA, Wootton M (2001) Changes in protein composition during grain development in wheat. Aust J Agric Res 52:485–493

Pei YH, Wang AL, An XL, Li XH, Zhang YZ, Huang XQ, Yan YM (2007) Characterization and comparative analysis of three low molecular weight glutenin C-subunit genes isolated from *Aegilops tauschii*. Can J Plant Sci 87:273–280

Petersen G, Seberg O, Yde M, Berthelsen K (2006) Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). Mol Phylogenet Evol 39:70–82

Pistón F, Martín A, Dorado G, Barro F (2004) Cloning and characterization of a gamma-3 hordein mRNA (cDNA) from *Hordeum chilense* (Roem. et Schult.). Theor Appl Genet 108:1359–1365

Pistón F, Martín A, Dorado G, Barro F (2005) Cloning and molecular characterization of B-hordeins from *Hordeum chilense* (Roem. et Schult.). Theor Appl Genet 111:551–560

Pistón F, Martín A, Dorado G, Barro F (2006) Cloning of nine r-gliadin mRNAs (cDNAs) from wheat and the molecular characterization of comparative transcript levels of g-gliadin subclasses. J Cereal Sci 43:120–128

Pistón F, León E, Lazzeri PA, Barro F (2008) Isolation of two storage protein promoters from *Hordeum chilense* and characterization

of their expression patterns in transgenic wheat. Euphytica 162:371–379

Sabelli PA, Shewry PR (1991) Characterization and organization of gene families at the *Gli-1* loci of bread and durum wheats by restriction fragment analysis. Theor Appl Genet 83:209–216

Schultz JF, Milpetz PB, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. Proc Natl Acad Sci USA 95:5857–5864

Shewry PR, Halford NG (2002) Cereal seed storage proteins: structures, properties and role in grain utilization. J Exp Bot 53:947–958

Wang ZQ, Long H, Zheng YL, Yan ZH, Wei YM, Lan XJ (2005) Cloning and analysis of LMW-GS genes from *Triticum aestivum* ssp. *Tibetanum Shao*. J Genet Genomics 32:86–93

Whelan JA, Russell NB, Whelan MA (2003) A method for the absolute quantification of cDNA using real-time PCR. J Immunol Methods 278:261–269

Wicker T, Yahiaoui N, Guyot R, Schalgenhauf E, Liu ZD, Dubcovsky J, Keller B (2003) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A^m genomes of wheat. Plant Cell 15:1186–1197

Yan Y, Hsam SLK, Yu JZ, Jiang Y, Zeller FJ (2003) Allelic variation of the HMW glutenin subunits in *Aegilops tauschii* accessions detected by Sodium Dodecyl Sulphate (SDS-PAGE), Acid Polyacrylamide Gel (A-PAGE) and Capillary Electrophoresis. Euphytica 130:377–385

Yan Y, Zheng J, Xiao Y, Yu J, Hu Y, Cai M, Li Y, Hsam SLK, Zeller FJ (2004) Identification and molecular characterization of a novel y-type *Glu-D^t1* glutenin gene of *Aegilops tauschii*. Theor Appl Genet 108:1349–1358

Zhang Y, Li Q, Yan Y, Zheng J, An X, Xiao Y, Wang A, Wang H, Hsam SLK, Zeller FJ (2006) Molecular characterization and phylogenetic analysis of a novel glutenin gene (*Dy10.1^t*) from *Aegilops tauschii*. Genome 49:735–745

Zhang YZ, Li XH, Wang AL, An XL, Zhang Q, Pei YH, Gao LY, Ma WJ, Appels R, Yan YM (2008) Novel x-type high-molecular-weight glutenin genes from *Aegilops tauschii* and their implications on the wheat origin and evolution mechanism of *Glu-D1-1* proteins. Genetics 178:23–33

Zhao H, Wang R, Guo A, Hu S, Sun G (2004) Development of primers specific for LMW-GS genes located on chromosome 1D and molecular characterization of a gene from *Glu-D3* complex locus in bread wheat. Hereditas 141:193–198

Zhao XL, Xia XC, He ZH, Gale KR, Lei ZS, Appels R, Ma W (2006) Characterization of three low-molecular-weight *Glu-D3* subunit genes in common wheat. Theor Appl Genet 113:1247–1259